

# **Infant communicative development assessed with the European Portuguese MacArthur–Bates Communicative Development Inventories short forms**

**Sónia Frota, Joseph Butler,  
Susana Correia and Cátia Severino**

Universidade de Lisboa, Portugal

**Selene Vicente**

Universidade do Porto, Portugal

**Marina Vigário**

Universidade de Lisboa, Portugal

First Language

1–21

© The Author(s) 2016



Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0142723716648867

[fla.sagepub.com](http://fla.sagepub.com)



## **Abstract**

This article describes the European Portuguese MacArthur–Bates Communicative Development Inventories short forms, the first published instruments for the assessment of language development in EP-learning infants and toddlers. Normative data from the EP population are presented, focusing on developmental trends for vocabulary learning, production of morphologically complex words and word combinations. Significant effects of gender were found for early word comprehension and production, as well as for toddlers' word production, word complexity and word combinations, with a consistent advantage for girls. By contrast, effects of socioeconomic status were limited to early word comprehension. A cross-language comparison with short form data for other languages was performed. Differences across languages only emerged for toddlers. However, effects of gender were found independently of language differences.

## **Corresponding author:**

Sónia Frota, Lisbon Baby Lab, Centro de Linguística, Faculdade de Letras, Universidade de Lisboa, Alameda da Universidade, 1600-214 Lisboa, Portugal.

Email: [sonia.frota@mail.telepac.pt](mailto:sonia.frota@mail.telepac.pt)

**Keywords**

Cross-linguistic comparison, gender differences, infants, language assessment, language development, Portuguese CDI, toddlers

Assessing early development of language skills is a challenging task, especially when the cooperation of the infant or toddler is required for a considerable amount of time, and when the tests administered demand the presence of highly trained examiners. In cases where language sampling can be used, data analysis is very time-consuming and also requires trained personnel. These factors impact on the sample sizes used, which tend to be small. One way of obtaining knowledge on early language development in a feasible way covering the range of variation in large samples is by using parental reports. This article describes the European Portuguese MacArthur–Bates Communicative Development Inventories short forms (EP-CDI SFs), and reports findings from a norming study of the EP population, focusing on developmental trends for vocabulary learning, production of morphologically complex words and word combinations in children aged 8–30 months.

The MacArthur–Bates Communicative Development Inventories (CDI; Fenson et al., 1993, 2007) are one of the best-known and widely used parental reports. The CDI was developed as a cost-efficient, reliable and valid means for assessing early language skills, and importantly for collecting data in large samples as required for establishing population-based norms. Initially developed for American English, the CDI has been adapted to more than 60 languages (Dale & Penfold, 2011), reflecting their linguistic and cultural differences while allowing for cross-linguistic comparisons (e.g., Bleses et al., 2008; Eriksson et al., 2012; Hamilton, Plunkett, & Schafer, 2000; Law & Roy, 2008; Wehberg et al., 2007). Evidence for the validity and reliability of the CDI has been reported in many studies for several languages (Fenson et al., 2007; Law & Roy, 2008 for reviews; Ring & Fenson, 2000 for English; Bornstein, Putnick, & De Houwer, 2006 for Dutch; O'Toole & Fletcher, 2010 for Irish; Trudeau & Sutton, 2011 for Quebec French).

However, the long forms of the CDI developed for American English and many other languages also showed limitations and restricted applicability and effectiveness in several research, educational and clinical settings. A major constraint is the considerable amount of time needed to complete the long form, together with demands on the literacy level of the parents. For example, the use of the CDI long form is problematic within research projects where parents and children are involved in many other procedures, in longitudinal studies that need repeated administration of the instrument, in studies with bilingual children which require the completion of CDIs for both languages, or in a clinical setting where a quick assessment of language ability is required, and where the education background of families varies. These limitations have motivated the development of short form versions of the CDI, which were shown to be at least as effective, reliable and valid as the long forms, and more applicable in the contexts described. Starting with Fenson, Pethick, Renda, and Cox (2000) for American English, short forms were developed for other languages, such as Spanish (Jackson-Maldonado, Marchman, & Fernald,

2013) and Galician (Pérez-Pereira & Resches, 2007). In all cases, the short forms showed equivalent developmental trends to the long forms.

To the best of our knowledge, the EP-CDI SFs are the first published instruments specifically designed for the assessment of early language skills and their development in European Portuguese-learning infants and toddlers. Unlike for other languages for which the CDI long forms were already available when the short forms were constructed, the development of the EP-CDI SFs was based on results from pilot studies and on databases of spontaneously produced child speech and child directed speech only, informed by prior knowledge of early language development in EP and of the language-specific patterns. The present article thus offers a new approach to the development of CDI short forms.

European Portuguese (EP) is a Romance language spoken by ca. 10,125,000 people in Portugal (INE, 2012),<sup>1</sup> with a range of dialectal variations that differentiate between varieties spoken in the north, in the center-southern regions, and the islands of Azores and Madeira (Segura, 2013). EP, similar to most other Romance languages, differs from English in its inflectional and word morphology, by marking gender and number on nouns, adjectives, pronouns and other nominal-like elements as well as on determiners, by showing a rich verbal inflection and by having prevailing derivation word-formation processes instead of compounding (Bauer, 1983; Vigário & Garcia, 2012). Spanish and Galician are languages closely related to EP, albeit with important differences at the segmental, prosodic and lexical levels. Unlike Spanish and Galician, but similarly to English, EP is characterized by pre- and post-tonic phonological vowel reduction (Mateus & Andrade, 2000; Vigário, 2003). EP differs from Spanish in the relative distributions of word stress patterns within the final three syllable window (Gibson, 2011; Vigário, Frota, & Martins, 2010). With respect to word shapes, the frequency of words larger than a binary foot (e.g., bigger than two syllables) is similar in EP and Spanish, and higher than in English, but the frequency of monosyllabic words is higher in EP than in Spanish (Roark & Demuth, 2000; Vigário, Freitas, & Frota, 2006). Furthermore, EP sound structure offers stronger cues for word segmentation than those available in other Romance languages, although weaker than in English (Vigário, 2003). These and other language-specific patterns may impact on the acquisition of words and related language development (Demuth, 2006; Millotte et al., 2010; Vigário et al., 2006).

A tool to assess early language skills and their development in European Portuguese-learning infants and toddlers was needed not only to provide large-scale studies of early language development and establish population-based norms, but also to meet the demands posed within an array of research, educational and clinical settings. Relevant examples are the need to assess infants' language skills prior to some experimental study, to evaluate potential early predictors of later (typical and atypical) language development in prospective studies, or to measure early language skills in atypical populations. The CDI instruments have been widely used for all of these purposes (Bergelson & Swingley, 2012; Friedrich, Herold, & Friederici, 2009; Law & Roy, 2008).

Charting the development of early language skills in EP-learning children contributes to the growing body of CDI-based research and its ongoing debates on developmental trends and their variation (within and between languages), the relationship between vocabulary comprehension and production and between vocabulary and word combinations, as well as the effects of socioeconomic status (SES) and gender on development.

Previous studies have reported globally similar developmental trends and individual range of variation across languages, with comprehension preceding production, an acceleration in vocabulary production in the second year of life and a strong correlation between vocabulary production and grammatical development (Fenson et al., 2007; Law & Roy, 2008). However, some languages have shown a different pace in development, e.g., Danish and French children exhibit a slower pace, and American English children show a faster pace in development (Bleses et al., 2008; Millotte et al., 2010; Wehberg et al., 2007). It has been hypothesized that at least some of these differences might be related to linguistic factors, namely to aspects of sound structure like strong reduction in Danish and absence of word boundary cues in French (but see Hamilton et al., 2000 for the potential role of social/cultural variables). As outlined above, EP has a mix of more Romance-like and more English-like sound features, being thus an interesting addition to cross-linguistic comparisons. Although in most CDI studies parents with lower SES are under-represented (Fenson et al., 2007; Fenson et al., 2000; Hamilton et al., 2000; Jackson-Maldonado et al., 2013; Simonsen et al., 2014), effects of SES on language development emerge clearly in some studies (Fenson et al., 2007; Jackson-Maldonado et al., 2013) and are absent in others (Hamilton et al., 2000; Pérez-Pereira & Resches, 2007). A further debate concerns gender differences in emerging language skills (Bornstein, Hahn, & Haynes, 2004; Eriksson et al., 2012; Lovas, 2011; Simonsen et al., 2014). Generally, an advantage for girls has been found, although not consistently for all language skills measured (e.g., vocabulary comprehension), all ages considered (e.g., younger infants) and all languages studied (e.g., Austrian and Galician infants).

The present study investigates early development of language skills in typically developing EP-learning monolinguals between 0;8 and 2;06, using the EP adaptation of the CDI short forms. First, we briefly describe the methodology followed in this adaptation and the subsequent norming study that was conducted. Then, we report on the reliability and validity of the EP-CDI SFs. Finally, we address three questions: (1) What characterizes the developmental trends in language skills in EP-learning infants and toddlers? (2) Whether and how SES and gender differences affect the development of these skills; and (3) How do EP developmental trends compare to those reported for American English, Spanish and Galician, identified from CDI short form data?

## Method

### *Development of the European Portuguese CDI short forms*

Work on the European Portuguese adaptation of the CDI started in 2011. Authorization by the CDI Advisory Board for such adaptation was granted and official approval of the European Portuguese CDI short forms (EP-CDI SFs) was obtained in 2012. The norming study was concluded in 2014.

The EP adaptation consists of one infant form (SFI), covering development between 8 and 18 months, and one toddler form (SFII) covering development between 16 and 30 months. The infant form assesses the size of receptive and productive vocabulary. The toddler form measures vocabulary production, production of complex words and word combinations. In the construction of the EP-CDI SFs, the guidelines for short forms

outlined in Fenson et al. (2000), information obtained from databases of spontaneously produced child speech and child directed speech based on longitudinal corpora, knowledge from prior work on the acquisition of EP and the language-specific patterns of the language, as well as parental feedback, were all taken into account. The inventories were developed in three phases, briefly described below.

The first step was based on a longer version with 350 vocabulary items adapted from the American long form. This version was tested in a pilot study with parental report data from 53 children between 1;4 and 2;6, and revealed several difficulties, namely the time spent on completing the form was too long, exceeding the time constraints of most parents and many parents struggled with the form, especially those with weaker reading skills. Based on this first pilot study, the decision was taken to construct short forms of the CDI.

The second step was an adaptation of the American short forms (Fenson et al., 2000), informed by the data and parental suggestions collected in the first pilot. This version of the EP-CDI SFs was tested in a pilot study, with 20 infant forms and 104 toddler forms collected. The data obtained were used to revise the EP-CDI SFs according to the following criteria: (A) meeting the general guidelines for short forms (varying age of acquisition, inclusion of early- as well as late-appearing words, avoidance of individual, regional and other kinds of biases, avoidance of ambiguous words, balance among the semantic and structural linguistic categories of the CDI); (B) considering results from databases of spontaneously produced child speech and child directed speech obtained from longitudinal corpora; (C) approximating the EP-specific patterns in what concerns word shapes (in number of syllables), stress pattern and syllable type distribution; (D) amending the instructions given in the forms to reflect parents' suggestions and questions.

Thus, in the third step, criteria A and B were met by combining data from the two pilot studies with spontaneous data from longitudinal child speech (PLEX5 – Frota et al., 2012) and child directed speech corpora (CDS-EP – Frota, Cruz, Martins, & Vigário, 2013).<sup>2</sup> Importantly, the inclusion of early- and late-appearing words in either the infant or the toddler EP-CDI SFs mitigated potential floor and ceiling effects, respectively. All possible biases and ambiguous words were eliminated from the vocabulary list. The revised SFs included items from all of the semantic CDI categories for the infant form (SFI) and the toddler form (SFII), with similar semantic and morphosyntactic distributions as in the American English originals, favoring the comparability across languages necessary for conducting cross-linguistic studies. This does not eliminate the need to reflect some language specificities. Due to the rich verbal inflection that characterizes EP, helping verbs do not have the same status as in English. On the other hand, word complexity is a prominent feature. As production studies of language acquisition in EP have indicated that morphological complexity in word formation seems to begin through the use of stressed-like suffixes such as *-zinho* (which added to a noun like *leão* 'lion' yields a new word meaning 'little lion' – Vigário & Garcia, 2012), one item of the 'helping verbs' category in SFII was replaced by an item to assess the production of complex words and its developmental pattern.

Item selection also took into consideration the EP-specific patterns (criterion C) that characterize word shapes, stress patterns and syllable type distribution, with the goal of approximating the SFs to the frequencies of use found in the language. Vigário et al.

(2006) reported that word shapes in child speech data are closer to adult speech than to CDS patterns. In the EP-CDI SFs, word shape distributions are more closely matched to the general pattern of the language found in adult speech, thus reflecting this finding, and similar trends hold for syllables types and stress patterns (Correia, 2010; PLEX5; CDS-EP; FrePoP – Frota, Vigário, Martins, & Cruz, 2010).<sup>3</sup> By and large, the most frequent patterns in the language overall (i.e., disyllabic words and penult stress) were boosted in the SFs.

Finally, the instructions given to parents/caregivers were amended (criterion D) to clarify any questions raised by the different possible pronunciations or formats of a given word.

The definitive EP-CDI SFs are one-page questionnaires that are easy and quick to administer. The EP-CDI SFI consisted of 90 vocabulary items with separate response columns for comprehension, and comprehension and production. The EP-CDI SFII focuses on expressive vocabulary and contains 99 vocabulary items, with the last one targeting complex word formation. Item 100 assesses the ability to produce word combinations as not yet present ('not yet'), emergent ('sometimes') or already well-developed ('often'). The EP-CDI SFs can be found at <http://labfon.letras.ulisboa.pt/babylab/pt/CDI/> (see also the dedicated website for further details).

### *Participants and procedure*

According to the National Statistics Institute (INE), there were 511,054 children between 0 and 4 years of age in Portugal in 2012. The information provided by INE stated the number of children, by gender, in the five areas of residence (North, Center, Lisbon, Alentejo, Algarve) in continental Portugal and the islands of Azores and Madeira. Considering this target population split into groups based on region and gender (7 regions × male, female), we computed our sample using G\*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), with the following parameters: medium effect size of .25, statistical significance of .05 and confidence interval of .95. Given the age ranges of SFI and SFII, we obtained a sample size of 407 and 429, respectively. Thus, in total 836 forms were considered for the norming study. Quota sampling was used (Castillo, 2009) to determine sample distribution by region and gender according to the INE population data.

Data were collected from all seven regions in collaboration with educational institutions, namely 84 nursery schools participated in the study. Schools from areas of diverse SES were included. During the entire process of administration of the EP-CDI SFs, members of the research team were in regular contact with the schools. Most forms were completed following this procedure. For the region of Lisbon, some forms were completed by caregivers visiting the baby lab to participate in other studies. Overall, 1507 forms were completed (637 infants and 870 toddlers), with 671 excluded from the final sample (due to unknown child age, other incomplete information, bilingual home environments). The normative sample was limited to children only exposed to European Portuguese as their native language, i.e., EP was the only language spoken in the child's home environment. Medical exclusion criteria included hearing loss and Down syndrome. Forms were not collected for children reported to have such medical conditions, based on the information provided by the nursery school teacher or the caregiver. An overview of the participants by month and gender is provided in the Appendix 1.



**Table 1.** Parental employment/educational status in the norming sample compared with the Portuguese population with children between 0 and 4 years of age (INE, 2012).

Parental employment status	CDI-I		CDI-II		Portugal
	N (407)	%	N (429)	%	%
Highly qualified	244	59.95	256	59.67	32.09
Medium qualified	117	28.75	141	32.87	27.26
Low qualified and workers	15	3.69	12	2.80	29.40
Unemployed	15	3.69	20	4.66	10.82
Missing	16	3.93	0	0.00	

The final sample of children was balanced for gender (51% boys, 49% girls) and the sibling status of the children matched that of the population with children (1 child over 50%, 2 children around 34% and 3 or more between 8 and 11%). Parental employment/educational status in the CDI sample, however, differs from the national numbers, with the majority of participating families in the highly or medium qualified categories and few in the low qualified category. – see Table 1.<sup>4</sup> These numbers probably reflect the sampling procedures and are in line with other CDI norming studies by showing a skewing towards higher educational levels of the caregivers (Fenson et al., 2007; Fenson et al., 2000; Jackson-Maldonado et al., 2013; Kristoffersen et al., 2012; Simonsen et al., 2014). Nevertheless, the EP-CDI sample included a fairly good range of employment/educational statuses.

### Data analysis

The data from the norming study were analyzed by means of descriptive statistics, inferential statistics and curve fitting, following Fenson et al. (2000) and Fenson et al. (2007). Mean scores on the short forms were computed as a function of age and gender. Analyses of variance (ANOVAs) were run to examine the effects of age, gender and SES on language outcomes. For each age in months, percentile scores were computed and fitted scores were calculated through growth-curve modeling using the logistic function. The advantages of this method to treat cross-sectional samples with variability, as well as the kinds of curves often found in growth processes, are described in Fenson et al. (2000) and Fenson et al. (2007). Normative percentile tables with fitted scores for boys and girls by month, for every 5th percentile level from the 5th to the 99th rank are found in the EP-CDI website.

## Results

### Reliability and validity

Reliability for the EP-CDI SFs was evaluated by computing Cronbach's coefficient alpha. The results approached 1.0: .99 for both the infant and toddler forms. These results indicate a high degree of internal consistency.

We also considered content and concurrent validity. With respect to content validity, the items included in the EP-CDI SFs were drawn from the developmental literature on EP and tested in earlier versions of the instruments. Furthermore, the items within the SFs closely approximate the EP-specific patterns in the relevant domains, namely word shape, word stress distribution and syllable type distribution. In this way, the content of the SFs relates both to the developmental skills the forms are designed to measure and to the specific patterns of the language being acquired.

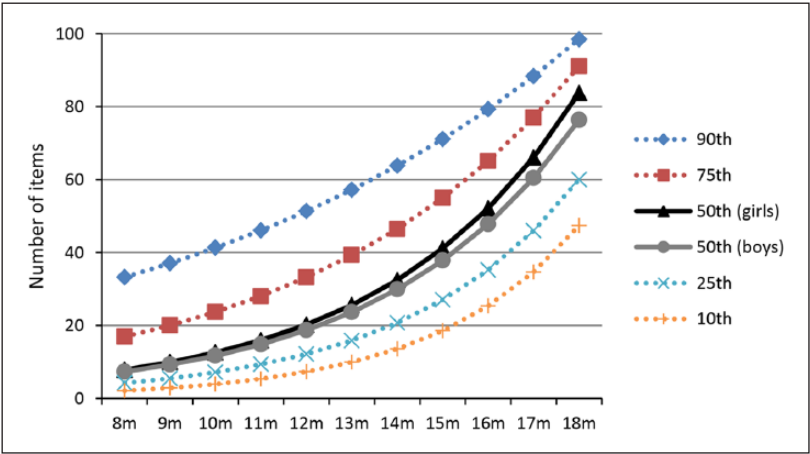
Concurrent validity was determined by assessing the relation between the CDI scores and child performance in spontaneous speech samples (along the lines of Kristoffersen et al., 2012). A comparison was made between the lexical items included in a lexicon based on spontaneously produced child speech (PLEX5 – Frota et al., 2012), and the words included in the EP-CDI SFs. For this comparison, a combined score from PLEX5 was considered including mean age of emergence, item production across children and overall frequency of occurrence. Each word was given a score from 3 to 0 based on these three factors (e.g., for emergence, early scored 3, late scored 1, intermediate scored 2 and missing scored 0). The maximum combined score for each item was 9 ( $3 \times 3$ ). This score was converted to a percentage, and this percentage was compared with the CDI scores, by means of Pearson correlations. For example, the lexical item *dar* ‘to give’, which is very frequent in child speech and emerges at 1;03, scored 100 in the spontaneous production data, and had a CDI score of 85.1 (SFI, comprehension), 46.2 (SFI, production) and 89.4 (SFII, production). The Pearson correlation results were as follows: for the infant form, .602 for vocabulary comprehension ( $p < .001$ ) and .694 for vocabulary production ( $p < .001$ ); for the toddler form, .744 for vocabulary production ( $p < .001$ ). These results indicate substantial correlations between child vocabulary in spontaneous speech samples and the CDI measures.

### *Developmental trends, gender and SES*

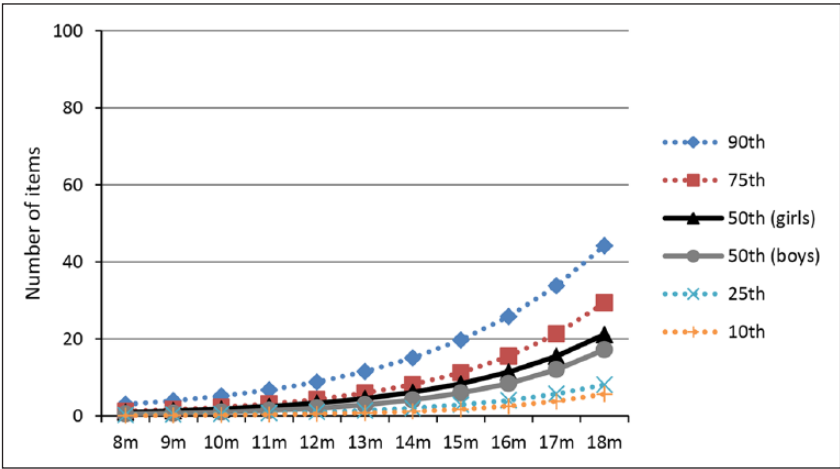
Results for developmental trends in communicative development are presented as growth curves based on fitted scores as a function of age in months for the 90th, 75th, 50th (median), 25th and 10th percentiles (along the lines of Fenson et al., 2007; Fenson et al., 2000; Jackson-Maldonado et al., 2013; see also the data analysis section above). The 50th percentile is shown separately for boys and girls. ANOVAs were calculated using raw scores to examine the impact of age and gender on language outcomes, as well as of SES. For some of these analyses, children were categorized into younger and older age groups (infant form: 8–12 and 13–18; toddler form: 16–20, 21–25 and 26–30 months) following previous studies (Fenson et al., 2007; Jackson-Maldonado et al., 2013). For more detailed descriptive statistics we refer to the EP-CDI website (<http://labfon.lettras.ulisboa.pt/babylab/pt/CDI/>). Finally, a cross-language comparison with short form data for American English, Spanish and Galician was performed.

*Developmental trends: Infants.* Developmental trends for vocabulary comprehension and production in infants by age and gender are given in Figures 1 and 2, respectively. The results support the well-established pattern in language acquisition literature that children’s receptive vocabulary extraordinarily exceeds their expressive one during the very first years of life. In addition, expressive vocabulary increased noticeably with age.





**Figure 1.** Words understood as a function of age (months), gender and percentile level. Fitted scores (infant short form, EP-CDI SFI).



**Figure 2.** Words produced as a function of age (months), gender and percentile level. Fitted scores (infant short form, EP-CDI SFI).

At age 8 months half of the infants did not understand more than 7 words. By 1;5, 50% of the infants are able to understand at least 55 words, whereas they produce no more than 11 words. Moreover, from 8 to 18 months, children make more progress in receptive (median of 7–82) than in expressive vocabulary (median of 0–28). Not surprisingly, individual differences are a constant in lexical development, as shown by the large range of scores within each month, with larger differences in receptive than in expressive vocabulary.

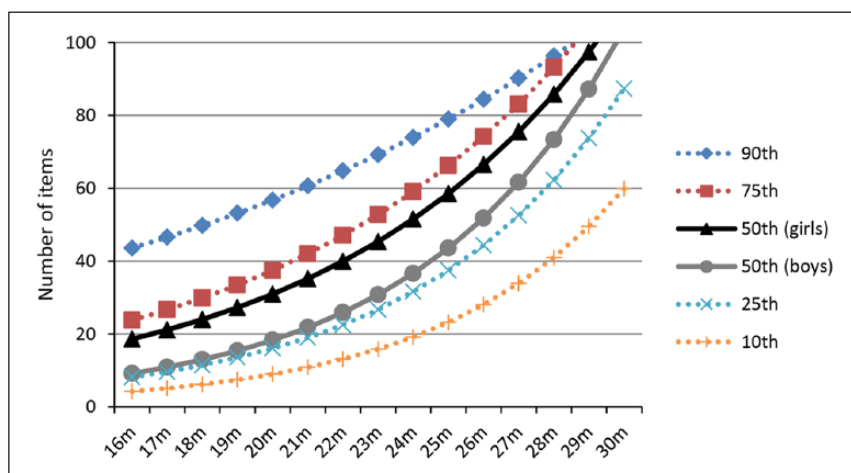
Word comprehension shows a steady increase with age across all percentile levels, with girls scoring higher than boys for all age groups. The median scores (50th) increased

to approximately double between 8 and 11 months, and they double again by 1;2. More than one-third of the infants scored at the maximum level at 18 months (90 words). The difference between medians of girls and boys ranged from +1 at 8 months to +8 at 1;6. A Gender (2)  $\times$  Age Group (2) between-subjects ANOVA on words understood yielded significant main effects of gender ( $F(1,406) = 4.16, p < .05, \eta_p^2 = .01$ ) and age group ( $F(1,406) = 222.02, p < .001, \eta_p^2 = .36$ ), with no interaction between the two factors ( $F < 1$ ). Overall, mean scores were higher for girls (8–12:  $M = 29.04, SD = 25.2$ ; 13–18:  $M = 64.99, SD = 25.41$ ) than for boys (8–12:  $M = 22.54, SD = 19.5$ ; 13–18:  $M = 61.27, SD = 26.86$ ). The effect of socioeconomic status (SES) was also examined. A 2 (Age Group)  $\times$  3 (SES; low, medium and highly qualified) ANOVA revealed the expected main effect of age group ( $F(1,375) = 59.77, p < .001, \eta_p^2 = .14$ ), and also a main effect of SES ( $F(2,375) = 7.45, p < .01, \eta_p^2 = .04$ ). No interactions between the two factors were found ( $F < 1$ ). Overall, children from medium qualified families understood more words ( $M = 58.14, SD = 32.69$ ) than children from lower or high levels of SES (low:  $M = 49.37, SD = 31.55$ ; high:  $M = 41.92, SD = 28.77$ ), the difference between highly and medium qualified being statistically significant ( $p < .001$ ).

As shown in Figure 2, the size of productive vocabulary in infants is quite small until 13 months, remaining near floor level and showing few differences in the percentile levels. After 13 months, the increase in production scores accelerates throughout the subsequent months, with a consistent advantage for girls. For example, at age 1;0 infants at the 90th percentile produce only 9 words (10% of the inventory), whereas at 1;6 these infants produce around 44 words, that represent 49% of the checklist. Half of the infants were reported to produce not more than 5 words at 1;0, whereas they produced at least 11 words at 1;4 and 28 at 1;6. The difference between medians of girls and boys ranged from +1 at 8 months to +4 at 1;6. A Gender (2)  $\times$  Age Group (2) ANOVA on words produced indicated significant main effects of gender ( $F(1,406) = 7.46, p < .01, \eta_p^2 = .02$ ) and age group ( $F(1,406) = 103.47, p < .001, \eta_p^2 = .2$ ), with no interaction between gender and age ( $F < 1$ ). Again, as expected, older children produced more words than younger children, and mean scores were higher for girls (8–12:  $M = 5.43, SD = 8.15$ ; 13–18:  $M = 19.05, SD = 17.68$ ) than for boys (8–12:  $M = 2.91, SD = 4.15$ ; 13–18:  $M = 14.74, SD = 13.25$ ). For vocabulary production, no effect of SES was observed ( $F(2,375) = 1.44, p = .24, \eta_p^2 = .01$ ), with no interaction between SES and Age Group ( $F < 1$ ).

The correlation between receptive and productive vocabulary in the EP-CDI SFI was .63 ( $p < .001$ ), dropping to .41 ( $p < .001$ ) after age was removed. This moderate correlation is comparable to the values reported by Fenson et al. (2000) for American English and Pérez-Pereira and Resches (2007) for Galician.

**Developmental trends: Toddlers.** Figure 3 provides the developmental trend for vocabulary production in toddlers by age and gender. During this period there is a substantial vocabulary growth, with the scores reflecting a gradual increase with age and a relatively stable variation across the age range. Between 18 and 22 months, vocabulary size nearly doubles (median 22 and 41, respectively). Ceiling effects were observed at 27 months at the top half of the distribution. Indeed, by this age a median child used more than 80 words of the 99 included in checklist. Girls consistently scored higher than boys for all age groups. The difference between medians of girls and boys shows a slower beginning

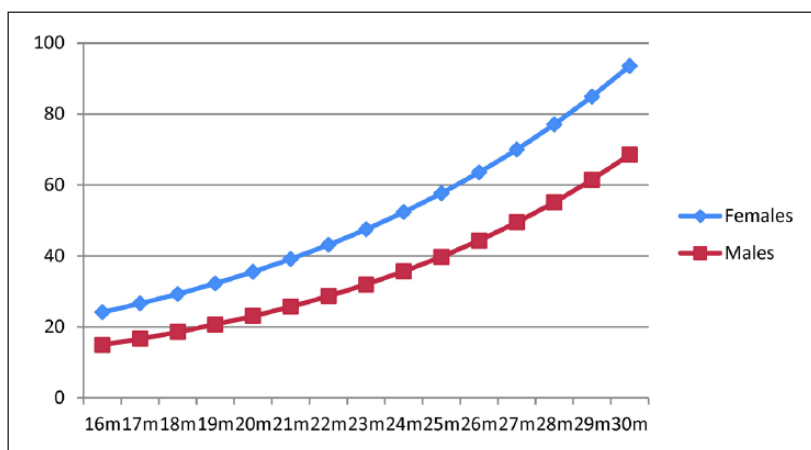


**Figure 3.** Words produced as a function of age (months), gender and percentile level. Fitted scores (toddler short form, EP-CDI SFII).

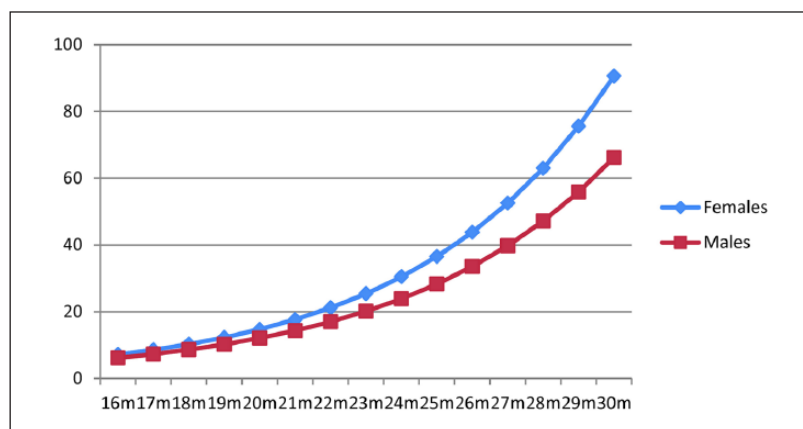
for boys as well as a slower growth. The differences ranged from +10 at 16 months to +15 at 25 months. After this age, the difference starts reducing, but girls still used 10 more words than boys at 29 months.

Not surprisingly, a Gender (2)  $\times$  Age Group (3) between-subjects ANOVA performed on words produced revealed significant main effects of gender ( $F(1,428) = 33.97, p < .001, \eta_p^2 = .07$ ) and age group ( $F(2,428) = 152.55, p < .001, \eta_p^2 = .42$ ), but no interaction between factors was observed ( $F < 1$ ). Overall, as observed in the infant form data, older children of both genders produced more words than children in the younger age groups, and mean scores showed a consistent advantage for girls at every age group (16–20:  $M = 35.33, SD = 22.11$ ; 21–25:  $M = 59.64, SD = 24.99$ ; 26–30:  $M = 85.10, SD = 14.87$ ) than for boys (16–20:  $M = 23.62, SD = 19.78$ ; 21–25:  $M = 43.23, SD = 24.96$ ; 26–30:  $M = 73.63, SD = 22.95$ ). Like for infants, no effect of SES on developmental trends in productive vocabulary was observed for toddlers ( $F < 1$ ).

Figure 4 presents the month-by-month trend for the production of complex words, by gender, based on the vocabulary item from EP-CDI SFII that assessed the production of morphologically complex words. Half of the children were reported to produce complex words by 26 months of age, and the percentage of complex word production rose from below 20% on average at 16 months to near 90% at 30 months, with girls showing a consistent advantage over boys at all age groups and a faster growth in the production of complex words. A Gender (2)  $\times$  Age Group (3) between-subjects ANOVA yielded significant main effects of gender ( $F(1,428) = 11.98, p < .01, \eta_p^2 = .03$ ) and age group ( $F(2,428) = 34.29, p < .001, \eta_p^2 = .14$ ), with no interaction ( $F < 1$ ). Pairwise comparisons revealed significant differences between all age groups ( $p < .001$ ; 16–20:  $M = 22, SD = 42$ ; 21–25:  $M = 52, SD = 50$ ; 26–30:  $M = 72, SD = 45$ ). Once again, mean scores were consistently higher for girls (with an overall mean of 57%) than for boys (with an overall mean of 40%). As for productive vocabulary, SES, unlike gender, had no effect on the production of complex words ( $F < 1$ ).



**Figure 4.** Proportion of children producing complex words (words ending with *-zinho*, e.g., *leãozinho* 'little lion') by age and gender (fitted data).



**Figure 5.** Proportion of children combining words 'often', by age and gender (fitted data).

A strong correlation was found between expressive vocabulary and production of complex words ( $r = .68, p < .001$ ).

Figure 5 presents the developmental trend for word combinations, by month and gender, based on the 'often' answer, that indicates an already well-developed ability to combine words.

There is a substantial increase with age of the percentage of children reported to be combining words, from low levels at 16 months, rising steadily up to 22 months and then with an acceleration to around 80% at 30 months. The acceleration in the frequency of word combinations is more evident for girls than for boys. The word combination data were further analyzed to evaluate the effects of age and gender by means of ANOVA. All the responses to the item assessing the ability to produce word

combinations were considered (i.e., ‘not yet’, ‘sometimes’ and ‘often’). A Gender (2)  $\times$  Age Group (3) between-subjects ANOVA yielded significant main effects of gender ( $F(1,404) = 7.16, p < .01, \eta_p^2 = .02$ ) and age group ( $F(2,404) = 86.1, p < .001, \eta_p^2 = .3$ ), with no significant interaction between the two factors ( $F(2,404) = 1.93, p = .15, \eta_p^2 = .01$ ).<sup>5</sup> For age group, pairwise comparisons revealed significant differences between all age groups ( $p < .001$ ; 16–20:  $M = .48, SD = .67$ ; 21–25:  $M = 1.1, SD = .75$ ; 26–30:  $M = 1.63, SD = .52$ ). Mean scores were higher for girls (with an overall mean of 1.17) than for boys (with an overall mean of .98). By contrast, a SES (3)  $\times$  Age Group ANOVA found no significant effect of SES ( $F(2,404) = 1.11, p = .33, \eta_p^2 = .01$ ), or interaction with age ( $F < 1$ ).

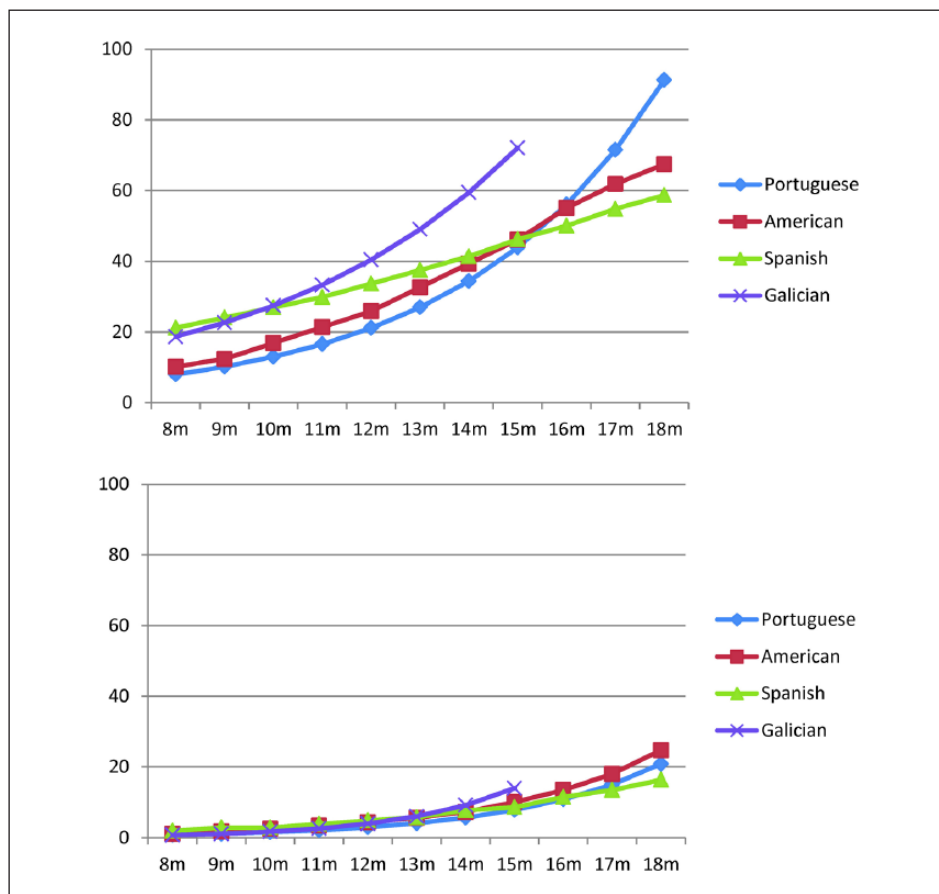
There is a strong correlation between the ability to combine words and the vocabulary score ( $r = .76, p < .001$ ), and a weaker correlation between word combinations and age ( $r = .59$ ). This finding points to the sensitivity of vocabulary growth as a measure of early syntactic development, which had already been reported for the CDI long form versions (e.g., Fenson et al., 2007), and thus strengthens the effectiveness of the short form versions of the CDI to detect early grammatical development patterns (as also highlighted in Fenson et al., 2000; Pérez-Pereira & Resches, 2007).

### Cross-language comparison

The CDI has been adapted and normed for different languages taking into account their linguistic and cultural differences, thus making comparisons across languages possible. These comparisons are especially suited when data collection instruments, sociodemographic samples and the analyses used are similar across studies. Although results from cross-linguistic comparisons need to be considered with caution given the potential number of candidate factors that may influence them, they are certainly a necessary step towards a better understanding of developmental trends in early language skills (Bleses et al., 2008; Wehberg et al., 2007). In this section, the fitted scores from the CDI short forms for American English (Fenson et al., 2000), Spanish (Jackson-Maldonado et al., 2013), Galician (Pérez-Pereira & Resches, 2007<sup>6</sup>) and European Portuguese are compared. The instruments and analyses used across these studies are similar, as well as the sociodemographic characteristics of the samples for English, Spanish and EP. For Galician, the sample shows two differences: a skewing towards lower educational level of the caregivers, and bilingualism is the norm. The present cross-language comparison is a step towards the general goal of contributing to the understanding of developmental trends in early language skills.

For the infant data, we compared scores for vocabulary comprehension and vocabulary production. The respective developmental trends for the four languages are plotted by month in Figure 6.

For comprehension, an Age Group (2 – 8–12 and 13–18 months)  $\times$  Language (4) ANOVA revealed the expected significant effect of age group ( $F(1,40) = 63.21, p < .001, \eta_p^2 = .66$ ), but no effect of language ( $F(3,40) = 1.35, p = .28, \eta_p^2 = .11$ ) and no interaction ( $F(3,40) = 1.12, p = .35, \eta_p^2 = .09$ ). Similarly, for vocabulary production a significant effect of age group was found ( $F(1,40) = 41.27, p < .001, \eta_p^2 = .56$ ), no effect of language ( $F(3,40) < 1$ ) and no interaction ( $F(3,40) < 1$ ). These findings indicate that overall developmental trends are similar across the four languages.

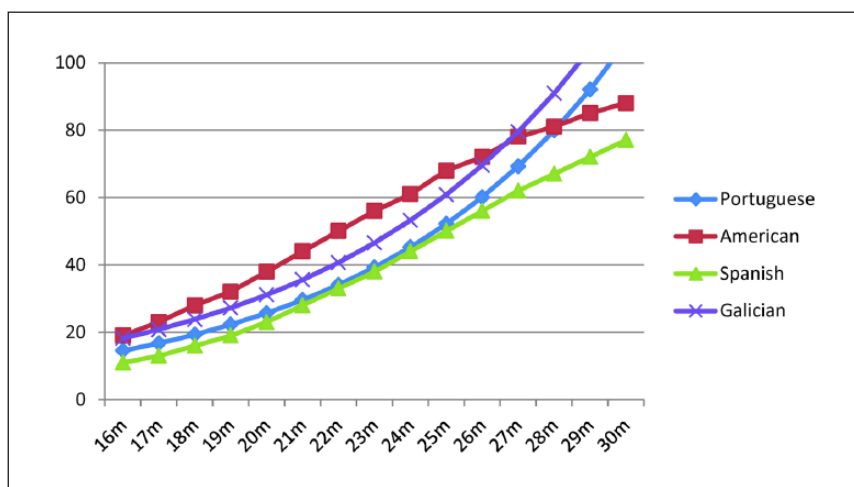


**Figure 6.** Infant form: Fitted vocabulary comprehension (top panel) and vocabulary production (bottom panel) scores by language (50th percentile).

The effect of gender was also examined across EP, English and Spanish. For comprehension, an Age Group (2)  $\times$  Gender (2)  $\times$  Language (3) ANOVA revealed a significant effect of age group ( $F(1,65) = 96.57, p < .001, \eta_p^2 = .64$ ), with no effects of gender ( $F(1,65) = 2.39, p = .13, \eta_p^2 = .04$ ) or language ( $F(2,65) < 1$ ), and no significant interactions ( $F(2,65) < 1$ ) with the exception of the borderline interaction between age group and language ( $F(2,65) = 2.96, p = .06, \eta_p^2 = .1$ ). This was explained by the overall lower scores of Spanish in the older age group, together with a larger difference between boys and girls in EP and American English, with girls scoring higher. For production, the ANOVA showed a significant effect of age group ( $F(1,65) = 65.3, p < .001, \eta_p^2 = .73$ ) and an almost significant effect of gender ( $F(1,65) = 3.84, p = .055, \eta_p^2 = .07$ ), but no effect of language ( $F(2,65) < 1$ ) and no significant interactions. Overall, girls (8.2) had higher scores than boys (5.93).

Figure 7 presents the month-by-month vocabulary production scores for toddlers across the four languages. An Age Group (3 – 16–20, 21–25 and 26–30 months)  $\times$





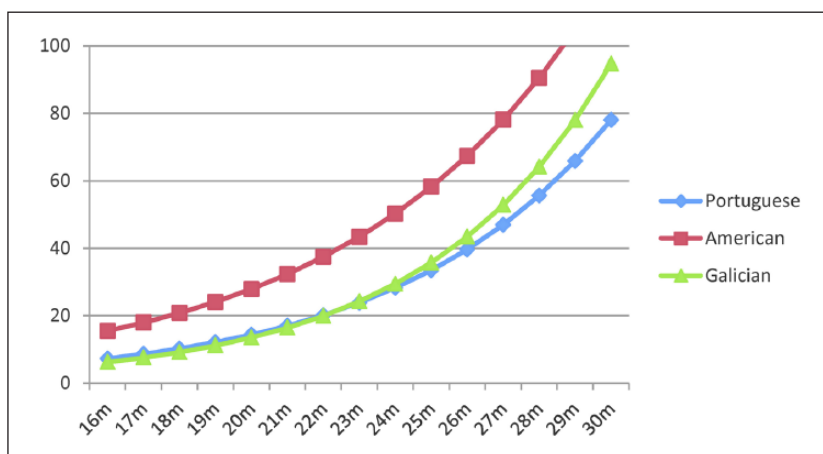
**Figure 7.** Toddler form: Fitted vocabulary production scores by language (50th percentile).

Language (4) ANOVA revealed a significant effect of age group ( $F(2,59) = 160.95$ ,  $p < .001$ ,  $\eta_p^2 = .87$ ) and language ( $F(3,59) = 6.58$ ,  $p < .01$ ,  $\eta_p^2 = .29$ ), with no interaction ( $F(6,59) = 1.36$ ,  $p = .25$ ,  $\eta_p^2 = .15$ ). For language, pairwise comparisons revealed significant differences between all language pairs except EP and Spanish, and English and Galician. Thus, unlike for the infant data, differences across languages are apparent in the toddler data.

The effect of gender was analyzed across EP, English and Spanish. An Age Group (3)  $\times$  Gender (2)  $\times$  Language (3) ANOVA revealed a significant effect of age group ( $F(2,89) = 248.83$ ,  $p < .001$ ,  $\eta_p^2 = .87$ ), language ( $F(2,89) = 15.35$ ,  $p < .001$ ,  $\eta_p^2 = .3$ ) and gender ( $F(1,89) = 29.17$ ,  $p < .001$ ,  $\eta_p^2 = .29$ ). Overall, girls (52.32) had higher scores than boys (41.47). There was a significant interaction between age group and language ( $F(4, 89) = 2.95$ ,  $p < .05$ ,  $\eta_p^2 = .14$ ), with no other interactions ( $F < 1$ ). The significant interaction was driven by Spanish boys scoring as high as EP ones in the two older age groups, unlike in the younger group. For language, pairwise comparisons showed significant differences between all language pairs. Thus, strong effects of language and gender were found for the toddler productive vocabulary data.

Finally, the scores for word combinations were compared across EP, Galician and English, using the fitted data for the 'often' response presented in Figure 8. An Age Group (3)  $\times$  Language (3) ANOVA revealed the expected effect of age group ( $F(2,44) = 89.67$ ,  $p < .001$ ,  $\eta_p^2 = .83$ ) and a significant effect of language ( $F(2,44) = 14.14$ ,  $p < .001$ ,  $\eta_p^2 = .44$ ), with no interaction ( $F(4,44) = 1.3$ ,  $p = .29$ ,  $\eta_p^2 = .13$ ). For language, pairwise comparisons revealed significant differences between all languages pairs ( $p < .001$ ) except EP and Galician. Therefore, like for vocabulary production, an effect of language was found for word combinations.

In sum, the results indicate that developmental trends are more similar across languages in the younger ages, whereas strong differences across languages may emerge in older ages, and a gender effect arises, which is stronger for production.



**Figure 8.** Toddler form: Fitted data for 'often' responses by language.

## Discussion

In the present article we described the European Portuguese CDI short forms and reported findings from a population-based study, focusing on developmental trends for vocabulary learning, production of morphologically complex words and word combinations in children aged 8–30 months. Three questions were addressed: (1) What characterizes the developmental trends in language skills in EP-learning infants and toddlers? (2) Whether and how SES and gender differences affect the development of these skills; and (3) How do EP developmental trends compare to those reported for American English, Spanish and Galician, identified from CDI short form data?

We found high internal consistency, comparable, for example, to that reported for the American English short forms (Fenson et al., 2000) or the Galician ones (Pérez-Pereira & Resches, 2007). Furthermore, we measured content and concurrent validity and found a good agreement between the content and results from the EP-CDI SFs and the developmental literature on EP. As the EP-CDI SFs were based on pilot studies and on data-bases of spontaneously produced child speech and child directed speech, informed by prior knowledge of early language development in EP and of its language-specific patterns, and not constructed from already available CDI long forms, we have shown that this new approach produces equally reliable and valid CDI data.

The findings for EP general trends in vocabulary development showed that, despite the larger individual differences, comprehension precedes production, and receptive vocabulary is characterized by a steady increase with age whereas expressive vocabulary shows an acceleration in vocabulary growth in the second year of life. These findings are in accordance with findings for other languages (Bleses et al., 2008; Fenson et al., 2007; Fenson et al., 2000; Hamilton et al., 2000; Jackson-Maldonado et al., 2013; Kern, 2007; Pérez-Pereira & Resches, 2007; Simonsen et al., 2014). For vocabulary production, ceiling effects were observed only after 27 months and exclusively at the top half of the distribution, similar to trends reported in short form data for American English, Spanish and Galician. This suggests that the EP-CDI SFs may also be appropriate for use with

at-risk or language-impaired children, who are expected to display lower scores even at older ages. More research is needed to explore the potential of the tool for clinical purposes, but preliminary data from two ongoing projects with at-risk children of autism and SLI and children with Down syndrome seem promising.

Word complexity increased with age, with half of the children reported to produce complex words by 26 months and 90% by 30 months. Complex word production was strongly correlated with expressive vocabulary. For word combinations, a substantial increase with age was also found, with the age of 22 months signaling a period of considerable growth. Overall, the trend observed is consistent with production data of longitudinal studies available for EP (Frota, Cruz, Matos, & Vigário, 2016). The reported age is also a key moment for the growth in word combinations in the short form data for American English and Galician (Fenson et al., 2000; Pérez-Pereira & Resches, 2007). Like in studies on other languages (e.g., Fenson et al., 2000; Law & Roy, 2008; Pérez-Pereira & Resches, 2007), there is a strong correlation between productive vocabulary and the ability to combine words. The fact that ceiling level is not reached at 30 months, both for word complexity and word combinations, suggests that the EP-CDI SFII may be useful with older children, in particular in cases of language delay or language impairment.

Gender differences were found to be statistically significant for all language skills measured, showing a consistent advantage for girls regardless of age. Although such an advantage has been reported for other languages, it seems to be modulated by the type of language skills mastered and the age range, with receptive vocabulary and different ages not showing gender effects or only displaying weak effects (Bornstein et al., 2004; Eriksson et al., 2012; Jackson-Maldonado et al., 2013; Lovas, 2011). In our findings, girls clearly outperformed boys, already at the infant stage and increasing with age. Thus, our results seem to lend support both to the biological differences in neurological maturity view (generalized effect in early stages) and to the gender socialization view (widening of gender effect over time – Lovas, 2011). Independently of the factors contributing to these effects, which require further study, our findings indicate that separate norms for girls and boys should be used in assessing early development of language skills.

By contrast, the effect of SES in our data was very limited (and restricted to vocabulary comprehension only), adding to the CDI literature where such effects are generally absent from developmental trends (e.g., Hamilton et al., 2000; Pérez-Pereira & Resches, 2007). However, since very low-SES families were under-represented in our study, the present findings should be considered with caution.

Finally, we compared EP developmental trends to those reported for American English, Spanish and Galician, identified from CDI short form data. A gender effect emerged across languages, even in infant production data, strengthening previous reports of robust gender differences regardless of language (Bornstein et al., 2004; Eriksson et al., 2012; Simonsen et al., 2014). Developmental trends were more similar across languages in infants, whereas strong differences across languages emerged in toddlers, both for vocabulary production and word combinations. Among the differences, European Portuguese paired with Spanish or with Galician, but not with American English, suggesting that in a language with mix sound properties like EP, the Romance-like features overcome the Germanic-like features, at least for the language abilities measured. Indeed, similar observations have already been made in the literature on the development of sound structure in EP (Frota et al., 2016; Vigário et al., 2006). These

findings are also in line with observations about differences in developmental trends between other languages, arguably due to language-specific sound features (Bleses et al., 2008; Millotte et al., 2010; Wehberg et al., 2007).

Although further research is needed to fully ascertain the applicability of the EP-CDI short forms for research, education and clinical settings, including typical and atypical populations, and also bilingual children, the present evidence strongly suggests that these are valuable and promising tools to assess early language development in EP in a broad range of contexts.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research received funding from the Foundation for Science and Technology, Portugal (projects PTDC/CLE-LIN/108722/2008 and EXCL/MHC-LIN/0688/2012).

### Notes

1. Another ca. 2,200,000 people are native speakers of EP who are emigrants in different parts of the world.
2. PLEX5 and CDS-EP are publicly available resources that can be accessed from <http://frepop.lettras.ulisboa.pt/>. These resources include mean age of emergence for word production and word frequency by CDI age range.
3. The FrePoP database is publicly available at <http://frepop.lettras.ulisboa.pt/>. We used the FrePoP spontaneous child speech sample from 1;0 to 3;0 (~70,000 words), and for adult speech a spontaneous speech sample ~500,000 words.
4. Professional categories followed the employment/educational status classification of the Portuguese Classification of Jobs (CNP, 2010, retrieved from <http://cdp.portodigital.pt/profissoes/classificacao-nacional-das-profissoes-cnp>), with 6 of the 9 CNP categories represented in the sample.
5. In some of the forms (24 in total), this item was left unanswered, and thus the *N* obtained for this item differs from the *N* for the vocabulary scores.
6. For Galician, the fitted data were not directly available in Pérez-Pereira and Resches (2007) and thus were calculated from the raw scores given in the paper. Also, for the infant form the data available range from 8 to 15 months. Gender data are not provided. For Spanish, word combination data are not available from Jackson-Maldonado et al. (2013).

### References

- Bauer, L. (1983). *English word-formation*. Cambridge, UK: Cambridge University Press.
- Bergelson, E., & Swingle, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, 109, 3253–3258.
- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, 35, 619–650.
- Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, 24, 267–304.

- Bornstein, M. H., Putnick, D. L., & De Houwer, A. (2006). Child vocabulary across the second year: Stability and continuity for reporter comparisons and a cumulative score. *First Language*, 26, 299–316.
- Castillo, J. (2009). *Quota sampling applied in research*. Retrieved from <http://explorable.com/quota-sampling>
- Correia, S. (2010). *The acquisition of primary word stress in European Portuguese* (Doctoral dissertation). University of Lisbon, Portugal.
- Dale, P., & Penfold, M. (2011). *Adaptations of the MacArthur–Bates CDI into non-U.S. English languages*. Retrieved from <http://mb-cdi.stanford.edu/documents/AdaptationsSurvey7-5-11Web.pdf>
- Demuth, K. (2006). Crosslinguistic perspectives on the development of prosodic words [Guest Editor, Special Issue]. *Language and Speech*, 49, 129–297.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., . . . Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Development Psychology*, 30, 326–343.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., . . . Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes Publishing.
- Fenson, L., Pethick, S., Renda, C., & Cox, J. L. (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21, 95–116.
- Friedrich, M., Herold, B., & Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex*, 45, 662–676.
- Frota, S., Correia, S., Severino, C., Cruz, M., Vigário, M., & Cortês, S. (2012). *PLEX5: A production lexicon of child speech for European Portuguese*. Lisboa, Portugal: Laboratório de Fonética (CLUL/FLUL).
- Frota, S., Cruz, M., Martins, F., & Vigário, M. (2013). *CDS\_EP: A lexicon of child directed speech from the FrePoP database (0;11 to 3;04)*. Lisboa, Portugal: Laboratório de Fonética (CLUL/FLUL).
- Frota, S., Cruz, M., Matos, N., & Vigário, M. (2016). Early prosodic development: Emerging intonation and phrasing in European Portuguese. In M. Armstrong, N. Henriksen, & M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 295–324). Amsterdam, The Netherlands: John Benjamins.
- Frota, S., Vigário, M., Martins, F., & Cruz, M. (2010). *FrePOP – Frequency patterns of phonological objects in Portuguese: Research and applications*. Lisboa, Portugal: Laboratório de Fonética (CLUL/FLUL). Retrieved from <http://frepop.letras.ulisboa.pt/>
- Gibson, M. (2011). A typology of stress in Spanish non-verbs. *Ianua: Revista Philologica Romanica*, 11, 1–30.
- Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language*, 27, 689–705.
- INE (2012). Censos 2011 – Resultados definitivos. Portugal. Lisboa, Portugal: Instituto Nacional de Estatística. Retrieved from [http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=ine\\_censos\\_publicacao\\_det&contexto=pu&PUBLICACOESpub\\_boui=73212469&PUBLICAC\\_OESmodo=2&selTab=tab1&pcensos=61969554](http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=ine_censos_publicacao_det&contexto=pu&PUBLICACOESpub_boui=73212469&PUBLICAC_OESmodo=2&selTab=tab1&pcensos=61969554)

- Jackson-Maldonado, D., Marchman, V., & Fernald, L. (2013). Short-form versions of the Spanish MacArthur–Bates Communicative Development Inventories. *Applied Psycholinguistics*, 34, 837–868.
- Kern, S. (2007). Lexicon development in French-speaking infants. *First Language*, 27, 227–250.
- Kristoffersen, K., Simonsen, H., Bleses, D., Wehberg, S., Jørgensen, R., Eiesland, E., & Henriksen, L. (2012). The use of the Internet in collecting CDI data: An example from Norway. *Journal of Child Language*, 40, 567–585.
- Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health*, 13, 198–206.
- Lovas, G. S. (2011). Gender and patterns of language development in mother–toddler and father–toddler dyads. *First Language*, 31, 83–108.
- Mateus, M. H., & Andrade, E. (2000). *The phonology of Portuguese*. Oxford, UK: Oxford University Press.
- Millotte, S., Morgan, J., Margules, S., Bernal, S., Dutat, M., & Christophe, A. (2010). Phrasal prosody constrains word segmentation in French 16-month-olds. *Journal of Portuguese Linguistics*, 9, 67–86.
- O'Toole, C., & Fletcher, P. (2010). Validity of a parent report instrument for Irish speaking toddlers. *First Language*, 30, 199–217.
- Pérez-Pereira, M., & Resches, M. (2007). Elaboración de las formas breves del Inventario do Desenvolvimento de Habilidades Comunicativas. Datos normativos y propiedades psicométricas [Development of the short forms of the Communicative Development Inventory. Normative data and psychometric properties]. *Infancia y Aprendizaje*, 30, 565–588.
- Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*, 20, 141–159.
- Roark, B., & Demuth, K. (2000). Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Conference on Language Development* (pp. 597–608). Somerville, MA: Cascadia Press.
- Segura, L. (2013). Variedades dialectais do Português Europeu [Dialect variations in European Portuguese]. In E. P. Raposo, M. B. do Nascimento, M. A. Mota, L. Segura, & A. Mendes (orgs.). *Gramática do Português* [Portuguese Grammar] (Vol. 1, pp. 85–142). Fundação Calouste Gulbenkian/CLUL.
- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, 34, 3–23.
- Trudeau, N., & Sutton, A. (2011). Expressive vocabulary and early grammar of 16 to 30-month old children acquiring Quebec French. *First language*, 31, 480–507.
- Vigário, M. (2003). *The prosodic word in European Portuguese*. Berlin, Germany: Mouton de Gruyter.
- Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of the prosodic word in European Portuguese. *Language and Speech*, 49, 175–203.
- Vigário, M., Frota, S., & Martins, F. (2010). A frequência que conta na aquisição da fonologia: tipos ou tokens [The frequency that counts in the acquisition of phonology: types or tokens]. In A. M. Brito, F. Silva, J. Veloso, & A. Fiéis (orgs.), *Textos Seleccionados do XXV Encontro Nacional da APL* [Selected texts, XXV National Meeting of the Portuguese Association of Linguistics] (pp. 749–767). Porto, Portugal: APL.
- Vigário, M., & Garcia, P. (2012). Palavras complexas na aquisição da morfologia do Português: estudo de caso [Complex words in the acquisition of the morphology of Portuguese: a case study]. In A. Costa, C. Flores, & N. Alexandre (Eds.), *Textos Seleccionados do XXVII Encontro*



*Nacional da APL* [Selected texts, XXVII National Meeting of the Portuguese Association of Linguistics] (pp. 604–624). Lisboa, Portugal: APL.

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., & Basbøll, H. (2007). Danish children's first words: Analysing longitudinal data based on monthly CDI parental reports. *First Language*, 27, 361–383.

## Appendix I

### Participants EP-CDI SFI.

Month	Boys	Girls	Total
	N	N	N
8	6	15	21
9	22	18	40
10	14	16	30
11	19	27	46
12	16	22	38
13	27	20	47
14	26	23	49
15	23	18	41
16	19	12	31
17	22	15	37
18	13	14	27

### Participants EP-CDI SFII.

Month	Boys	Girls	Total
	N	N	N
16	3	3	6
17	7	5	12
18	7	9	16
19	14	19	33
20	11	9	20
21	20	18	38
22	28	16	44
23	15	20	35
24	17	18	35
25	14	16	30
26	28	15	43
27	11	18	29
28	19	19	38
29	18	16	34
30	7	9	16